DOI:
10.1088/1742-5468/2016/04/043306

Large deviations of spread measures for Gaussian matrices

J. Stat. Mech. (2016) 043306

(http://iopscience.iop.org/1742-5468/2016/4/043306)

OPEN ACCESS

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Disordered systems, classical and quantum

Large deviations of spread measures for Gaussian matrices

Fabio Deelan Cunden[1] **and Pierpaolo Vivo**[2]

[1] School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK

[2] Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK

E-mail: fabiodeelan.cunden@bristol.ac.uk

Received 4 February 2016
Accepted for publication 14 March 2016
Published 28 April 2016

CrossMark

Online at stacks.iop.org/JSTAT/2016/043306
doi:10.1088/1742-5468/2016/04/043306

Abstract. For a large $n \times m$ Gaussian matrix, we compute the joint statistics, including large deviation tails, of generalized and total variance—the scaled log-determinant H and trace T of the corresponding $n \times n$ covariance matrix. Using a Coulomb gas technique, we find that the Laplace transform of their joint distribution $\mathcal{P}_n(h, t)$ decays for large n, m (with $c = m/n \geqslant 1$ fixed) as $\hat{\mathcal{P}}_n(s, w) \approx \exp(-\beta n^2 J(s, w))$, where β is the Dyson index of the ensemble and $J(s, w)$ is a β-independent large deviation function, which we compute exactly for any c. The corresponding large deviation functions in real space are worked out and checked with extensive numerical simulations. The results are complemented with a finite n, m treatment based on the Laguerre–Selberg integral. The statistics of atypically small log-determinants is shown to be driven by the split-off of the *smallest* eigenvalue, leading to an abrupt change in the large deviation speed.

Keywords: random matrix theory and extensions, statistical inference

1742-5468/16/043306+20$33.00

Contents

1. Introduction

The standard deviation σ of an array of m data X_i is the simplest measure of how *spread* these numbers are around their average value $\bar{X} = (1/m)\sum_{i=1}^{m} X_i$. Suppose that the X_i's represent the final 'Physics' marks of m students of a high-school. Most worrisome scenarios for the headmaster would be a low $\bar{X}$ and/or a high σ, signaling an overall poor and/or highly non-uniform performance.

What if 'Physics' and 'Arts' marks are collected together? Detecting performance issues now immediately becomes a much harder task, as data may fluctuate together and in different directions. A two-dimensional scatter plot may help, though. The 'centre' of the cloud gives a rough indication of how well the students perform on average in both subjects. But how to tell in which subject the gap between excellent and mediocre students is more pronounced, or whether outstanding students in one subject also excel in the other?

In figure 1 (Bottom) we sketch two scatter plots of marks adjusted to have zero mean. A meaningful spread indicator seems to be the shape of the ellipse enclosing each cloud. For example, an almost circular cloud—like School 1—represents a rather uninformative situation, where your 'Arts' marks tell nothing about your 'Physics' skills, and vice versa. Conversely, a rather elongated shape—like School 2—highlights correlations between each student's marks in different subjects.

For a bunch of many scattered points it would be desirable to summarize the overall spread around the mean just by a single scalar quantity, like the *perimeter* or *area* of the enclosing ellipse. Not surprisingly, however, these indicators (taken individually) have evident shortcomings [30]. Surely a wiser choice is to combine more than a single

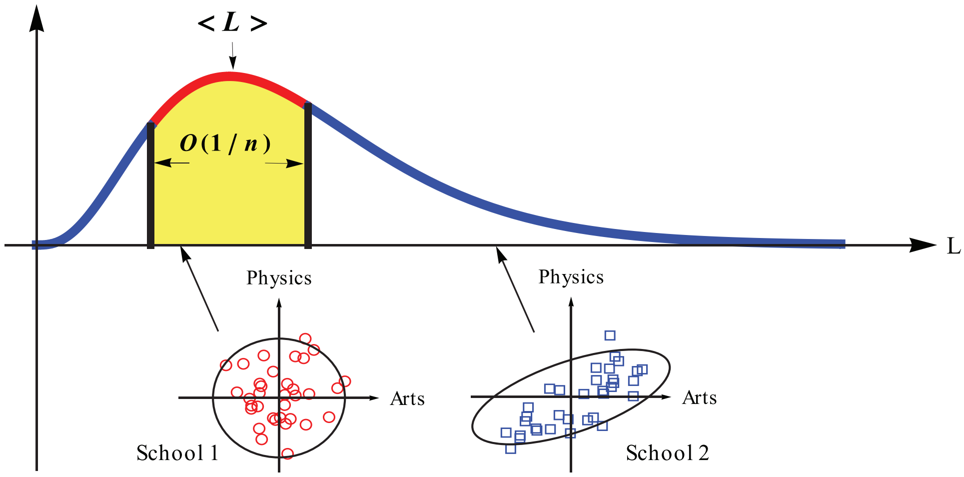

Figure 1. Top: sketch of the probability density of the likelihood ratio L of a Gaussian iid data set. In yellow, the typical region around the mean of order $\mathcal{O}(1/n)$. Larger fluctuations are referred to as atypical large deviations. Bottom: sketch of two multivariate data sets with $n = 2$ and $m = 35$. Each point represents a student, for two different schools, and his/her marks in Arts and Physics. The two datasets have same generalized variance H, but different total variance T. The likelihood ratio L of school 1 is compatible with the iid hypothesis, while the value of L for School 2 is atypically far from the average $\langle L\rangle$.

spread measure (like perimeter *or* area *alone*), to obtain a more revealing indicator. These issues arise naturally in multivariate statistics, and more mathematical tools and techniques are required compared to the univariate setting.

In this work, we compute the *joint* statistics of 'perimeter' and 'area' enclosing clouds of random high-dimensional data. Why this is a crucial (and so far unavailable) ingredient for an accurate data analysis will become clearer very shortly.

In the more general setting of n subjects and m students, their marks can be arranged in a $n\times m$ matrix $\mathcal{X}$, adjusted to have zero-mean rows. We then construct the normalized $n\times n$ covariance data matrix $\mathcal{S} = (1/n)\mathcal{X}\mathcal{X}^\dagger$, with non-negative eigenvalues $(\lambda_1, \ldots, \lambda_n)$, which is precisely the multi-dimensional analogue of the variance σ^2 for a single array. The surface and volume ('perimeter' and 'area' in the two-dimensional example) of the enclosing ellipsoid are related to the scaled trace and determinant of $\mathcal{S}$:

$$T = \frac{1}{n}\mathrm{Tr}\mathcal{S} \quad \text{and} \quad G = \det \mathcal{S}^{1/n}. \tag{1}$$

In statistics, these objects are called *total* and *generalized* variance respectively [2]. As discussed before, blending both estimators together would be preferable, like in the widely used positive scalar combination

$$L = T - H - 1, \tag{2}$$

called *likelihood ratio* [2], where H is the log-determinant of $\mathcal{S}$

$$H = \ln G = \frac{1}{n}\mathrm{Tr}\ln\mathcal{S}. \tag{3}$$

Values of L for different shapes of the data cloud are sketched in figure 1 (Bottom).

Now, suppose that we wish to test the hypothesis that the data X_{ij} (yielding a certain *empirical* L) are independent and identically distributed. What if an *atypically* high or low L (with respect to a null i.i.d. model) comes out from the data? We would be tempted to reject the test hypothesis outright. However, this might lead to a misjudgment, as atypical values of L for the null model *can* (and *do*) occur (just very rarely). What is the probability of this rare event? Here we provide a solution to this problem, computing the joint statistics of total and generalized variance for a large Gaussian dataset.

The derivation of these results relies on techniques borrowed from statistical mechanics and random matrix theory (RMT). We express the large deviation functions of spread indicators as excess free energies of an associated 2D Coulomb gas, whose thermodynamic limit is analyzed in the mean-field approximation valid for $n, m \to \infty$ with $m/n > 1$ fixed. This approach is complemented with a finite n, m analysis based on the 'Laguerre' version of the celebrated Selberg integral. The marriage between these two techniques provides an elegant solution to a challenging problem. In addition, our unifying framework recovers and extends some partial results earned by statisticians via other techniques.

The article is organized as follows. In section 2 we introduce the notation and we summarize our main results. We then elaborate at length on its consequences. Finally, we briefly discuss the relation of our findings with earlier works. Section 3 contains the derivations. First we summarize the 'Coulomb gas method' and we present a quite general algorithm to find the large deviation functions of linear statistics on random matrices (section 3.1). Then, in section 3.2 we turn to the actual proof. In section 4 we discuss two issues that are not captured by the Coulomb gas method. Finally we conclude with a summary and some open questions in section 5.

2. Setting and formulation of the results

We consider an ensemble of $n \times m$ matrices $\mathcal{X}$ whose entries are real, complex or quaternion independent standard Gaussian variables[3], labeled by Dyson's index $\beta = 1, 2$ and 4 respectively, and we form the $n \times n$ (real, complex or quaternion) sample covariance matrix

$$\mathcal{S} = \frac{1}{n} \mathcal{X}\mathcal{X}^{\dagger}. \tag{4}$$

This ensemble of random covariance matrices (positive semi-definite by construction) is known as the Wishart ensemble [55] with rectangularity parameter $c = m/n \geqslant 1$. Remarkably, in the Gaussian case, the joint probability density $\mathcal{P}(\lambda_1, \ldots, \lambda_n)$ of the positive $\mathcal{O}(1)$ eigenvalues of $\mathcal{S}$ is known explicitly [2, 25]

$$\mathcal{P}(\lambda_1, \ldots, \lambda_n) = \frac{1}{\mathcal{Z}_n} \mathrm{e}^{-\beta E[\boldsymbol{\lambda}]}, \quad E[\boldsymbol{\lambda}] = -\frac{1}{2} \sum_{i \neq j} \ln|\lambda_i - \lambda_j| + n \sum_k V(\lambda_k), \tag{5}$$

[3] The assumption of independence is not restrictive. If the entries of $\mathcal{X}$ are centered correlated Gaussian variables with positive definite covariance matrix Σ our methods can be applied to the matrix $\Sigma^{-1/2}\mathcal{S}\Sigma^{-1/2}$.

J. Stat. Mech. (2016) 043306

where the *energy* function $E[\boldsymbol{\lambda}]$ contains the external potential

$$V(\lambda)=\begin{cases}\dfrac{\lambda}{2}-\alpha\ln\lambda & \text{for } \lambda>0 \text{ if } \alpha>0\,(\text{or } \lambda\geqslant 0 \text{ if } \alpha=0)\\ +\infty & \text{otherwise}\end{cases}\quad \text{with } \alpha=\frac{c-1}{2}+\frac{1}{2n}-\frac{1}{\beta n}. \tag{6}$$

The normalization constant $\mathcal{Z}_n=\int \mathrm{e}^{-\beta E[\boldsymbol{\lambda}]}\mathrm{d}\boldsymbol{\lambda}$ is also known for any finite n from the celebrated Selberg integral [3, 26, 45]. The joint law of the eigenvalues (5) is the Gibbs–Boltzmann canonical distribution of a 2D Coulomb gas (logarithmic repulsion) constrained to stay on the positive half-line and subject to the external potential V at inverse temperature β (we adopt the usual physical convention that probabilities are zero in regions of infinite energy). As we shall see, the derivation of our result is independent of the restriction $\beta=1,2$ or 4. Therefore, from now on we shall consider non-quantized values[4] $\beta>0$.

We consider the scaled log-determinant H and trace T of the covariance matrix $\mathcal{S}$. In terms of the eigenvalues $\lambda_1,\dots,\lambda_n$, they read

$$H=\frac{1}{n}\sum_{i=1}^{n}\ln\lambda_i \quad \text{and} \quad T=\frac{1}{n}\sum_{i=1}^{n}\lambda_i. \tag{7}$$

Their joint probability law and Laplace transform are denoted respectively by

$$\mathcal{P}_n(h,t)=\langle\delta(h-H)\delta(t-T)\rangle,\quad \widehat{\mathcal{P}}_n(s,w)=\left\langle \mathrm{e}^{-\beta n^2(sH+wT)}\right\rangle, \tag{8}$$

where the average is taken with respect to the canonical distribution of the eigenvalues (5). Here we are interested in the large n behavior of $\mathcal{P}_n(h,t)$ and $\widehat{\mathcal{P}}_n(s,w)$ at logarithmic scales. More precisely, we show that for large n

$$\mathcal{P}_n(h,t)\approx \mathrm{e}^{-\beta n^2\Psi(h,t)} \qquad \text{and} \qquad \widehat{\mathcal{P}}_n(s,w)\approx \mathrm{e}^{-\beta n^2 J(s,w)}, \tag{9}$$

where $a_n\approx b_n$ stands for $\ln a_n/\ln b_n\to 1$ as $n\to\infty$.

The functions $\Psi(h,t)$ and $J(s,w)$ are called *rate function* and *cumulant generating function* (GF) respectively [21, 51]. It is a standard result in large deviation theory that the functions $\Psi(h,t)$ and $J(s,w)$ in (9) are related via a Legendre–Fenchel transformation.

Here we compute explicitly, for all $\beta>0$ and $c=m/n\geqslant 1$, the cumulant GF

$$J(s,w)=-\lim_{n\to\infty}\frac{1}{\beta n^2}\ln\widehat{\mathcal{P}}_n(s,w). \tag{10}$$

From now on we shall denote $s_c=(c-1)/2$. The main results of the paper are as follows.

A (Joint large deviation function of generalized and total variances). *Let $(\lambda_1,\dots,\lambda_n)$ be distributed according to (5)–(6) and let H and T as in (7). Their joint cumulant generating function $J(s,w)$ defined by (10) exists for $s\leqslant s_c$ and $w>-1/2$ and is given by*

$$J(s,w)=J_H(s)+J_T(w)-s\ln(1+2w), \tag{11}$$

[4] Eigenvalues obeying the Wishart statistics with general $\beta>0$ can be generated efficiently using Dumitru–Edelman tridiagonal construction [19].

where $J_T(w)$ *and* $J_H(s)$ *are the individual GF of cumulants of* T *and* H. *They are given explicitly by*

$$J_H(s) = -\lim_{n\to\infty}\frac{1}{\beta n^2}\ln\widehat{\mathcal{P}}_n(s,0) = \phi(s-s_c)-\phi(-s_c), \tag{12}$$

$$J_T(w) = -\lim_{n\to\infty}\frac{1}{\beta n^2}\ln\widehat{\mathcal{P}}_n(0,w) = \frac{c}{2}\ln(1+2w), \tag{13}$$

with $\phi(x) = -\frac{3}{2}x + x^2\ln(-2x) - \frac{(1-2x)^2}{4}\ln(1-2x)$ *for* $x\leqslant 0$.

We first discuss some consequences of this result. The derivation is postponed to section 3.

Remark 1. The large deviation functions $J(s, w)$, $J_H(s)$ and $J_T(w)$ are independent of β. This property is standard for 2D Coulomb gas systems.

Remark 2. The joint cumulant GF is not the sum of the single generating functions: $J(s,w)\neq J_H(s)+J_T(w)$ (T and H are not independent for large n).

Remark 3. For $c>1$ the GF is analytic at $s=w=0$ and the joint cumulants of T and H are obtained by evaluating the derivatives of $J(s, w)$ at $(s, w)=(0, 0)$. More precisely, to leading order in n for $\kappa,\ell\geqslant 0$

$$C_{\kappa,\ell}(H,T) = (-\beta n^2)^{1-(\kappa+\ell)}\frac{\partial^{\kappa+\ell}}{\partial^\kappa s\,\partial^\ell w}J(s,w)\Big|_{s=w=0}. \tag{14}$$

Note that $J(0, 0)=0$. Extracting the first cumulants, we obtain to leading order in n

$$\langle T\rangle = c, \qquad \langle H\rangle = -1-(c-1)\ln(c-1)+c\ln c, \tag{15}$$

$$\frac{\mathrm{var}(T)}{\omega_\beta(n,2)} = c, \quad \frac{\mathrm{var}(H)}{\omega_\beta(n,2)} = \ln\frac{c}{c-1}, \quad \frac{\mathrm{cov}(T,H)}{\omega_\beta(n,2)} = 1, \tag{16}$$

where we set $\omega_\beta(n,\ell)=(2/\beta n^2)^{\ell-1}$. The correlation coefficient $\left(\mathrm{cov}(H,T)/\sqrt{\mathrm{var}(T)\mathrm{var}(H)}\right) = 1/\sqrt{c\ln(c/(c-1))}$, independent of β, is positive for all values of c (if the 'area' increases, typically so does the 'perimeter'). Notice that the expression of $\mathrm{var}(H)$ does not cover the case $c=1$ (square data matrices). This case will be treated separately in section 4.

The decay of the higher order mixed cumulants $C_{\kappa,\ell}(H,T)$ for $\kappa+\ell>2$ is given to leading order in n by

$$C_{\kappa,\ell}(H,T) = \omega_\beta(n,\kappa+\ell)\{(\kappa-3)!\,[(1-c)^{2-\kappa}-(-c)^{2-\kappa}]\,\delta_{\ell,0} + c\,(\ell-1)!\delta_{\kappa,0} + (\ell-1)!\delta_{\kappa,1}\}. \tag{17}$$

Remark 4. The marginal probability densities $\mathcal{P}_H(h)=\langle\delta(h-H)\rangle$ and $\mathcal{P}_T(t)=\langle\delta(t-T)\rangle$ behave as

$$\mathcal{P}_H(h)\approx \mathrm{e}^{-\beta n^2\Psi_H(h)} \quad \text{and} \quad \mathcal{P}_T(t)\approx \mathrm{e}^{-\beta n^2\Psi_T(t)}, \tag{18}$$

J. Stat. Mech. (2016) 043306

where $\Psi_H(h)$ and $\Psi_T(t)$ are the individual rate functions of T and H. These individual rate functions should also be in principle computable as inverse Legendre–Fenchel transform of (12) and (13). However, for the scaled log-determinant H this is only possible for 'not too small' values ($h > -1$); this point will be discussed in more details in section 4. The expression of $\Psi_T(t)$ in the full range is instead remarkably simple

$$\Psi_T(t) = \frac{t-c}{2} + \frac{c}{2}\ln\left(\frac{c}{t}\right), \qquad (t > 0). \tag{19}$$

This analytic function is strictly convex and positive and it attains its unique minimum (zero) at $t = c$ (the asymptotic mean value of T, see (15)). This rate function provides information on the large n full probability density of T. We can identify three regimes:

(i) typical fluctuations of order $\mathcal{O}(1/n)$ about the average are described by the quadratic behavior of $\Psi_T(t)$ around its minimum at $t = c$, corresponding to asymptotically Gaussian fluctuations with mean and variance as in (15) and (16);

(ii) large deviations for $t \gg c$ (atypically *large* 'perimeters') exhibit an exponential decay (independent of the rectangularity parameter c);

(iii) for $t \ll c$ (atypically *small* 'perimeters') we find a c-dependent power law.

Summarizing:

$$\mathcal{P}_T(t) \approx \mathrm{e}^{-\beta n^2 \Psi_T(t)} \sim \begin{cases} t^{\beta n^2 c/2}, & (t \to 0) \\ \mathrm{e}^{-\beta n^2 \frac{(t-c)^2}{4c}}, & (t \sim c) \\ \mathrm{e}^{-\beta n^2 t/2}, & (t \to +\infty). \end{cases} \tag{20}$$

These predictions have been confirmed by extensive numerical simulations. A sample size of about $N = 10^8$ spectra of complex ($\beta = 2$) Wishart matrices has been efficiently generated using a tridiagonal construction [47]. The data are plotted in figure 2 and show a very good agreement with the behavior in (18).

Once the joint large-n behavior of generalized and total variances is known, one may easily derive a large deviation principle for any continuous function of them. For instance, from $\widehat{\mathcal{P}}(s, \omega)$, it is easy to compute the Laplace transform of the likelihood ratio $L = T - H - 1$ as $\widehat{\mathcal{P}}_L(s) = \langle \mathrm{e}^{-\beta n^2 s L}\rangle = e^{\beta n^2 s}\widehat{\mathcal{P}}(-s, s)$. Hence we have the following result.

B (Large deviations of the likelihood ratio). *The likelihood ratio cumulant GF is given by*

$$J_L(s) = -\lim_{n\to\infty} \frac{1}{\beta n^2} \ln \widehat{\mathcal{P}}_L(s) = J(-s, s) - s \qquad \text{for} \quad -1/2 < s \leqslant s_c, \tag{21}$$

with J as in (11). With the same notation as above, the cumulants of L at leading order in n follow by differentiations

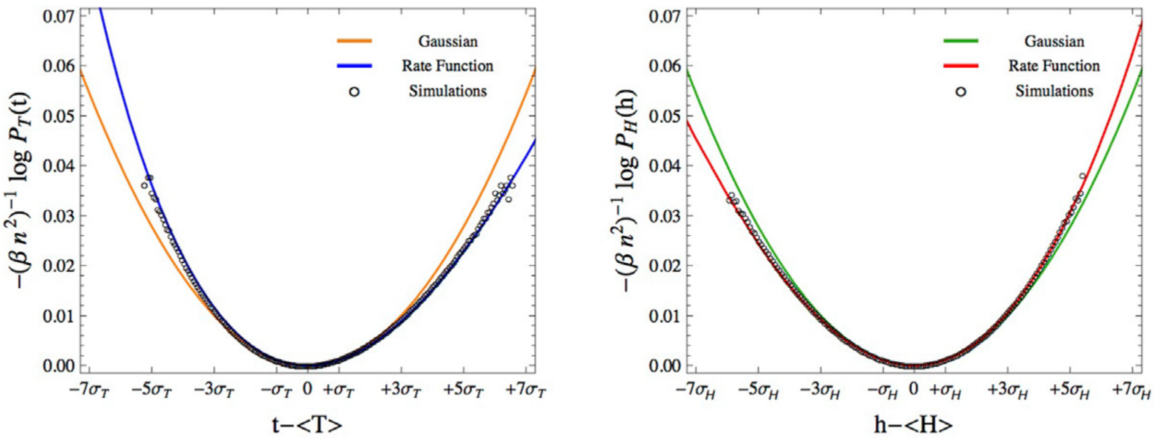

Figure 2. Numerical simulations (black circles) of complex ($\beta = 2$) Wishart matrices $\mathcal{S}$ of size $n = 15$ with $c = 2$. Here the sample size is $N = 2.5 \cdot 10^8$. Left: the numerical values (black circles) for the total variance $T = n^{-1}\sum_i \lambda_i$. The Gaussian approximation (orange line) with average $\langle T \rangle$ (15) and standard deviation $\sigma_T = \sqrt{\mathrm{var}(T)}$ (16) fits well the data for fluctuations of order $\sim 3\sigma_T$ but deviates strongly for atypical fluctuations. The global behavior is captured instead by the large deviation function (blue line) $\Psi_T(t)$ of (19). Right: numerical values (black circles) for the log-determinant $H = n^{-1}\sum_i \ln \lambda_i$. Again, the Gaussian approximation (green line) with average $\langle H \rangle$ (15) and standard deviation $\sigma_H = \sqrt{\mathrm{var}(H)}$ (16) describes well the data for fluctuations of order $\sim 3\sigma_H$ but deviates for larger fluctuations. The large deviation function (red line) $\Psi_H(h)$ of (45) provides a global description of the data. The critical point $h = -1$ (below which the large deviations change speed from n^2 to n) is not visible in the picture (for $n = 15$ and $c = 2$ the critical point is at $\sim 25\sigma_H$ to the left of $\langle H \rangle$).

$$C_\ell(L) = C_\ell(T) + (-1)^\ell C_\ell(H) + \delta C_\ell \quad \text{with} \quad \delta C_\ell = \omega_\beta(n,\ell)\frac{\ell\,!}{(1-\ell)}\theta(\ell-1) - \delta_{\ell,1}, \tag{22}$$

(θ is the Heaviside step function) for $\ell \geqslant 1$. This corresponds to typical fluctuations on a region $\mathcal{O}(1/n)$ around the mean

$$\langle L \rangle = c + (c-1)\ln(c-1) - c\ln c, \tag{23}$$

with variance[5]

$$\mathrm{var}(L) = \omega_\beta(n,2)\,[c + \ln(c/(c-1)) - 2]\,. \tag{24}$$

Note that, since T and H are not independent, the cumulants (22) of L involve the extra term δC_ℓ.

From result A, extracting the asymptotics of the first moments of T and H for $c \gg 1$ we recover classical results in multivariate analysis, valid when the sample size m is much larger than the number of variates n.

[5] Again, these results are not valid for $c = 1$.

J. Stat. Mech. (2016) 043306

C (Classical statistics). *In the regime $m \gg n \gg 1$, T and H become asymptotically Gaussian. More precisely, as $c \to \infty$*

$$\sqrt{\frac{\beta n^2}{2c}}\,(T-c) \to \mathcal{N}(0,1), \qquad \text{and} \qquad \sqrt{\frac{\beta c n^2}{2}}\,(H-\ln c) \to \mathcal{N}(0,1), \tag{25}$$

in distribution, where $\mathcal{N}(0,1)$ denotes a standard Gaussian variable.

To conclude this introductory section, we remark that our findings reproduce some known results for the typical fluctuations (mean and variance) of T, H and L separately, in the real case ($\beta = 1$) [2, 4, 29, 31, 32]. Moreover, the variances and covariances (16) and (24) can be computed for generic β using covariance formulae valid for one-cut β-ensembles of random matrices [6, 12].

A precious tool in classical statistics is the *Barlett decomposition* [5], which is useful to transform functions of strongly correlated eigenvalues of Wishart matrices (see (5)) into functions of independent (but not identical) chi-squared random variables. In the asymptotic regime $m \gg n$ this decomposition becomes sufficiently manageable to derive some interesting results. For real matrices, the limits (25) agree with classical theorems based on the Barlett decomposition (see e.g. [40]). From the results on H, the statistical behavior of the scaled determinant $G = \mathrm{e}^H$ can be easily derived. For statistics of determinants of random matrices (more general than the sample covariance matrices considered here), see [10, 33, 43, 49]. For more details on the classical methods in multivariate analysis we refer to the classical books [30, 40] and the excellent review [31] on the applications of RMT in multivariate statistics.

3. Derivation

We now turn to the derivation of result A. Results B and C follow as corollaries and hence their proof will be omitted. In section 3.1 we set up the variational problem in the framework of the 2D Coulomb gas thermodynamics. The Coulomb gas analogy for spectra of random matrix ensembles goes back to the seminal works by Wigner [54] and Dyson [20]. In particular, it was Dyson who first used this analogy to compute large random matrix statistics. This idea has been developed later and used in several areas of physics [8, 9, 11, 13–18, 22, 23, 28, 34–37, 50, 53]. In section 3.2 we solve the saddle-point equations and we compute explicitly $J(s, w)$, thus proving result A.

3.1. 2D Coulomb gas problem

The Coulomb gas calculation goes as follows. First we observe from (5) and (6) that the joint Laplace transform (8) is finite for $s \leqslant \alpha = s_c + \mathcal{O}(n^{-1})$ and $w > -1/2$. From (5) and (7)–(8), this Laplace transform can be written as the ratio of two partition functions

$$\widehat{\mathcal{P}}_n(s, w) = [\mathcal{Z}_n(s, w)/\mathcal{Z}_n(0,0)], \tag{26}$$

J. Stat. Mech. (2016) 043306

$$\mathcal{Z}_n(s,\, w) = \int \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_n \, \mathrm{e}^{-\beta E[\boldsymbol{\lambda};\, s,\, w]}. \tag{27}$$

$\mathcal{Z}_n(s,\, w)$ is the partition function of a constrained Coulomb gas, where the energy function $E[\boldsymbol{\lambda};\, s,\, w] = E[\boldsymbol{\lambda}] + \sum_k U_{s,w}(\lambda_k)$ contains now the additional single-particle potential

$$U_{s,\,w}(\lambda) = s \ln \lambda + w\lambda. \tag{28}$$

Note that $\mathcal{Z}_n(0, 0)$ is the partition function of the unconstrained gas and therefore it coincides with the normalization constant in (5).

Hence, the computation of the joint cumulant GF $J(s,\, w)$ amounts to evaluating the leading order in n of the partition function $\mathcal{Z}_n(s,\, w)$. More precisely, from (26) and (27), $J(s,\, w)$ may be expressed as the *excess free energy*

$$J(s,\, w) = -\lim_{n\to\infty} \frac{1}{\beta n^2} \left[\ln \mathcal{Z}_n(s,\, w) - \ln \mathcal{Z}_n(0, 0)\right] \tag{29}$$

of the Coulomb gas in the *effective potential* $V(\lambda) + U_{s,w}(\lambda)$ with respect to the unperturbed Coulomb gas ($s = w = 0$). This effective potential is bounded from below for $s \leqslant \alpha = s_c + \mathcal{O}(n^{-1})$ and $w > -1/2$ (the domain of existence of the Laplace transform $\widehat{\mathcal{P}}_n(s,\, w)$, of course). For any finite n, m, the excess free energy (29) can be computed exactly in terms of a Laguerre–Selberg integral [45] (see section 4). How to deal with the limit of large n? We show now how the task of computing $J(s,\, w)$ can be reduced to a variational problem.

First, we introduce the normalized density of the gas particles $\rho_n(\lambda) = n^{-1}\sum_{i=1}^{n} \delta(\lambda - \lambda_i)$, in terms of which any sum function on the eigenvalues $\lambda_1, \ldots, \lambda_n$ can be easily expressed. For instance, the log-determinant and trace (7), both linear statistics on $\mathcal{S}$, are conveniently expressed as linear functionals on $\rho_n(\lambda)$ as

$$H = \int \mathrm{d}\rho_n(\lambda)\, \ln \lambda \qquad \text{and} \qquad T = \int \mathrm{d}\rho_n(\lambda)\, \lambda. \tag{30}$$

Second, for large n, the energy function $E[\boldsymbol{\lambda};\, s,\, w]$ of the 2D Coulomb gas can be converted into a *mean-field energy functional* $E[\boldsymbol{\lambda};\, s,\, w] \sim n^2 \mathcal{E}[\rho_n;\, s,\, w]$, where

$$\mathcal{E}[\rho;\, s,\, w] = -\frac{1}{2} \iint_{\lambda \neq \lambda'} \mathrm{d}\rho(\lambda)\, \mathrm{d}\rho(\lambda')\, \ln|\lambda - \lambda'| + \int \mathrm{d}\rho(\lambda)\, V(\lambda) + \int \mathrm{d}\rho(\lambda)\, U_{s,w}(\lambda). \tag{31}$$

The mean-field functional (31) has been intensely studied in several fields. We refer to [16, 44, 46] for a detailed exposition and collection of known results. In particular, it is known that for large n the partition function $\mathcal{Z}_n(s,\, w)$ is dominated by $\rho_{s,\,w}(\lambda)$, the unique minimizer of the mean-field energy functional $\mathcal{E}[\rho;\, s,\, w]$ in the space of normalized densities:

$$\mathcal{Z}_n(s,\, w) \approx \exp(-\beta n^2 \mathcal{E}[\rho_{s,\,w};\, s,\, w]\,) \quad \text{with} \quad \mathcal{E}[\rho_{s,\,w};\, s,\, w] = \min_{\substack{\rho \geqslant 0 \\ \int \mathrm{d}\rho = 1}} \mathcal{E}[\rho;\, s,\, w]. \tag{32}$$

The meaning of the saddle-point density is the following: $\rho_{s,\,w}(\lambda)$ is the typical configuration of the eigenvalues yielding a prescribed value of log-determinant and trace

J. Stat. Mech. (2016) 043306

$$h(s,\,w) = \int \mathrm{d}\rho_{s,\,w}(\lambda)\,\ln\lambda \qquad t(s,\,w) = \int \mathrm{d}\rho_{s,\,w}(\lambda)\,\lambda. \tag{33}$$

Hence, a possible route to evaluate $J(s,\,w)$ consists of finding for all s, w the saddle-point density $\rho_{s,\,w}(\lambda)$ and inserting it back into the energy functional (31) to evaluate the leading order of $\mathcal{Z}_n(s,\,w)$ as in (32). This technique has been exploited in the last decade in many physical problems, mainly to compute the large deviations of *single* observables. However, this route entails the explicit computation of the mean-field energy (31) at the saddle-point density, which is not necessarily an easy task. The situation gets even worse in the case of joint statistics.

In certain situations one can use a shortcut (see [16]) based on a thermodynamic identity that has been stated rigorously in the language of large deviation theory [51] by Gärtner [27] and Ellis [21]. It is known that, if a cumulant GF $J(\vec{s})$ is differentiable in the interior of its domain, then the rate function $\Psi(\vec{x})$ is the Legendre–Fenchel transform of the cumulant GF (and hence, $J(\vec{s})$ is the inverse Legendre–Fenchel transform of $\Psi(\vec{x})$). This relation between rate function and cumulant GF can be exploited in our problem as follows (for a general mathematical discussion we refer to [16]). We assume first that $J(s,\,w)$ is differentiable. Therefore, the Gärtner–Ellis theorem ensures that $\Psi(h,\,t)$ is also smooth and given from $J(s,\,w)$ by the Legendre–Fenchel transformation

$$\Psi(h,\,t) = \sup_{s,\,w}\,[J(s,\,w) - (sh + wt)]. \tag{34}$$

The identity (34) can be written in the (almost) symmetric form

$$J(s,\,w) - \Psi(h,\,t) = sh + wt. \tag{35}$$

This equation should be interpreted with care. Indeed, in (35), there are only *two* independent variables, for instance s and w *or* h and t. The relation between the conjugate variables $(h,\,t)$ and $(s,\,w)$ is provided by

$$\frac{\partial J(s,\,w)}{\partial s} = h(s,\,w)\,, \frac{\partial J(s,\,w)}{\partial w} = t(s,\,w) \quad \text{or} \quad \text{equivalently} \quad \frac{\partial \Psi(h,\,t)}{\partial h} = s(h,\,t)\,, \frac{\partial \Psi(h,\,t)}{\partial h} = w(h,\,t), \tag{36}$$

where $h(s,\,w)$ and $t(s,\,w)$ are given in (7) and $s(h,\,t)$, $w(h,\,t)$ are the corresponding inverse maps. Hence, we can write the differential relations

$$\mathrm{d}J(s,\,w) = h(s,\,w)\mathrm{d}s + t(s,\,w)\mathrm{d}w \tag{37}$$

$$-\mathrm{d}\Psi(h,\,t) = s(h,\,t)\mathrm{d}h + w(h,\,t)\mathrm{d}t, \tag{38}$$

supplemented with the normalization condition $J(0,\;0) = 0$ (and hence $\Psi(h(0,0),$ $t(0,0)) = 0$). The expressions (37) and (38) can be interpreted as Maxwell relations among thermodynamic potentials, in our case the Helmholtz free energy and the enthalpy. However it is somewhat astonishing that these relations have not been applied in the Coulomb gas computations until very recently (for applications of (37) and (38) in physical models see [13, 14] and also [28]).

Using (37) and (38) one can use the following shortcut to compute the large deviations functions (for a detailed exposition we refer to [16]). For instance, in order to compute

$J(s,\, w)$ we only need to find the saddle-point density $\rho_{s,w}(\lambda)$ and compute $h(s,\, w)$ and $t(s,\, w)$ from (7). Then, $J(s,\, w)$ follows from integration of (37)

$$J(s,\, w) = \int_{(0,0)}^{(s,w)} \mathrm{d}J(s,\, w). \tag{39}$$

This shortened route of the Coulomb gas method provides an effective tool to evaluate large deviations functions. A large amount of unnecessary computations can be avoided and the task of computing *joint* large deviations becomes feasible. In the next section, we will use this strategy to derive our main result.

3.2. Saddle-point equation and large deviation functions

The first problem to overcome is to find the saddle-point density $\rho_{s,w}$ of the mean-field functional $\mathcal{E}[\rho;\, s,\, w]$. From (31), the stationarity condition of $\rho_{s,w}$ reads

$$\int \mathrm{d}\rho_{s,w}(\lambda') \ln|\lambda - \lambda'| - V(\lambda) - U_{s,w}(\lambda) = \text{const}, \qquad \text{for } \lambda \in \operatorname{supp}\rho_{s,w}, \tag{40}$$

where $\operatorname{supp}\rho_{s,w}$ denote the support of $\rho_{s,w}$ (for $\lambda \notin \operatorname{supp}\rho_{s,w}$ the left hand side is greater than or equal to the same constant). The physical meaning of (40) is clear: at equilibrium, the 2D Coulomb gas arranges itself in such a way that each particle has equal electrostatic energy (the left hand side of (40)).

Taking one further derivative with respect to λ, the resulting singular integral equation can be solved for $\rho_{s,w}(\lambda)$ using a theorem due to Tricomi ([52, section 4.3]), and the result reads

$$\rho_{s,w}(\lambda) = \frac{1+2w}{2\pi\lambda}\sqrt{(\lambda-\lambda_-)(\lambda_+-\lambda)}\ \mathbf{1}_{\lambda\in(\lambda_-,\lambda_+)}, \tag{41}$$

where the edges $\lambda_\pm$ of the support depend on s and w as

$$\lambda_\pm(s,\, w) = \frac{1}{1+2w}\left(1 \pm \sqrt{1-2(s-s_c)}\right)^2. \tag{42}$$

For $s = w = 0$ the density of the unperturbed gas $\rho_{0,0}(\lambda)$ coincides with the Marčenko–Pastur distribution [38] with edges $(1 \pm \sqrt{c})^2$. For $s < s_c$ the saddle-point density is bounded while at $s = s_c$ the lower edge $\lambda_-(s_c,\, w) = 0$ reaches the origin and $\rho_{s,w}$ acquires an inverse square root divergence there.

From $\rho_{s,w}$ the corresponding values of scaled log-determinant H and trace T are

$$\begin{cases} h(s,\, w) = \displaystyle\int \mathrm{d}\rho_{s,w}(\lambda)\, \ln\lambda = \varphi(s-s_c) - \ln(1+2w) \\ t(s,\, w) = \displaystyle\int \mathrm{d}\rho_{s,w}(\lambda)\, \lambda = \frac{c-2s}{1+2w}, \end{cases} \tag{43}$$

with $\varphi(x) = -1 + (1-2x)\ln(1-2x) + 2x\ln(-2x)$ for $x \leqslant 0$ (hence $h(s,\, w)$ is defined for $s \leqslant s_c$). Combining (43) and (39) we obtain the cumulant GF $J(s,\, w)$ as in result A.

In principle, one may also compute the rate function $\Psi(h, t)$ in the same way. However, it is not possible to write down $s(h,\, t)$ and $w(h,\, t)$ (the inverse maps of (43)) in terms of elementary functions. For simplicity, however, we show how to carry out the explicit computation for trace and log-determinant separately and establish the large n decay as in remark 4. Setting $s = 0$ we find $t(0, w) = c/(1 + 2w)$ from (43), and we immediately get the rate function of the scaled traces from (39) by integrating $w(t)$ (the inverse of $t(0, w)$)

$$\Psi_T(t) = -\int_{t(0,0)}^{t} w(t')\mathrm{d}t' = \int_c^t \frac{1}{2}\left(1 - \frac{c}{t'}\right)\mathrm{d}t' = \frac{t-c}{2} + \frac{c}{2}\ln\left(\frac{c}{t}\right). \tag{44}$$

This proves (19). Similarly, for the log-determinant H we have

$$\Psi_H(h) = -\int_{h(0,0)}^{h} s(h')\mathrm{d}h', \tag{45}$$

valid for $h > -1$ (see discussion in the next section), where $s(h)$ is the inverse of $h(s, 0)$.

4. Further results and discussion

The treatment in the previous section does not cover the following two issues:

- The case $c = 1$ (square data matrices), for which the leading term of the variance of $H = \ln\det\mathcal{S}$ (computed from the approach described above) is not defined (see equation (16)). What is the origin of this hitch?
- The origin of the condition $h > -1$ for the validity of the rate function $\Psi_H(h)$ in (45) seems mysterious. What is the mechanism governing the statistics of 'anomalously small' log-determinants, then?

We discuss these two issues in detail here.

4.1. The case $c = 1$ (square data matrices)

As already disclosed, if $\mathcal{S}$ is a Wishart matrix with $c = 1$ ($m = n$) the limiting variance of $H = \ln\det\mathcal{S}$ is not described by our large deviations result (see equation (16)). The origin of this hitch is as follows. Recall that the cumulant GF $J_H(s)$ is defined for $s \leqslant s_c = (c-1)/2$. Hence, for $c = 1$ (i.e. $m = n$) $J_H(s)$ is non-analytic in $s = s_c = 0$ and the cumulants cannot be obtained by differentiation.

A way to circumvent this problem is to *first* compute var(H) for *finite* n, and then evaluate its large n asymptotics. The joint Laplace transform $\widehat{\mathcal{P}}_n(s, w)$ of H and T can be indeed evaluated exactly also at finite n, m, using the Laguerre–Selberg integral [3, 26, 45]:

$$\frac{1}{n!}\int_0^{+\infty}\cdots\int_0^{+\infty}\prod_{i<j}|x_i - x_j|^{2p}\prod_{i=1}^{n} x_i^{q-1}\mathrm{e}^{-x_i}\mathrm{d}x_i = \prod_{j=0}^{n-1}\frac{\Gamma(q+jp)\Gamma((j+1)p)}{\Gamma(p)}, \qquad p, q > 0. \tag{46}$$

Using this identity, one may evaluate the Laplace transform $\widehat{\mathcal{P}}_n(s, w)$ as

$$\widehat{\mathcal{P}}_n(s,w)=\left(\frac{(\beta n/2)^s}{(2w+1)^{\frac{1}{2}+(\alpha-s)+\frac{1}{2n}}}\right)^{\beta n^2}\prod_{j=0}^{n-1}\frac{\Gamma\Big(\frac{\beta}{2}(j+2n(\alpha-s))+1\Big)}{\Gamma\Big(\frac{\beta}{2}(j+2n\alpha)+1\Big)},\tag{47}$$

with α as in (6). We have verified that our large n formulae reproduce with good accuracy the finite n, m result even for moderate values of n.

From (47) it is possible to extract the large deviation functions for the scaled trace T (this corresponds to $s=0$). On the other hand, the asymptotic in the variable s is not trivial.

However we can use this exact result to deduce var(H) for symmetric data matrices ($c=1$), the case that was not covered by our result A. Setting $c=1$ and $w=0$ in (47), we denote by $\widehat{\mathcal{P}}_n(s)\equiv\widehat{\mathcal{P}}_n(s,0)=\langle \mathrm{e}^{-\beta n^2 sH}\rangle$ the Laplace transform of H at finite $n=m$. We can compute the derivatives of $\widehat{\mathcal{P}}_n(s)$ as

$$\widehat{P}'_n(s)=\beta n^2\ln\left(\frac{\beta n}{2}\right)\widehat{P}_n(s)-\beta n\widehat{P}_n(s)\sum_{j=0}^{n-1}\psi_0\left(\frac{\beta}{2}(j+1-2ns)\right),\tag{48}$$

$$\begin{aligned}\widehat{P}''_n(s)=\beta n^2\ln\left(\frac{\beta n}{2}\right)\widehat{P}'_n(s)-\beta n\widehat{P}'_n(s)\sum_{j=0}^{n-1}\psi_0\left(\frac{\beta}{2}(j+1-2ns)\right)\\ +(\beta n)^2\widehat{P}_n(s)\sum_{j=0}^{n-1}\psi_1\left(\frac{\beta}{2}(j+1-2ns)\right),\end{aligned}\tag{49}$$

where $\psi_m(z)=\partial_z^{m+1}\ln\Gamma(z)$ is the m-Polygamma function [1]. In principle one can compute higher derivatives recursively, and evaluate the asymptotic values of the cumulants of H. For instance, average and variance of H are related to the derivatives $\widehat{P}'_n(s)$ and $\widehat{P}''_n(s)$ at $s=0$. Using the normalization $\widehat{P}_n(0)=1$ we get

$$\langle H\rangle=-\frac{1}{\beta n^2}\widehat{P}'_n(0)=-\ln\left(\frac{\beta n}{2}\right)+\frac{1}{n}\sum_{j=0}^{n-1}\psi_0\left(\frac{\beta}{2}(j+1)\right).\tag{50}$$

Using the Euler–Maclaurin summation formula [1] $\sum_{k=0}^{N}F(a+hk)=\frac{1}{h}\int_a^b\mathrm{d}t\,F(t)+\frac{1}{2}[F(b)+F(a)]+\dots$ (with $b=a+hN$), and the classical asymptotic $\ln\Gamma(az+b)\sim\ln\left(\sqrt{2\pi}\right)-az+\left(az+b-\frac{1}{2}\right)\ln(az)$, valid for $z\to\infty$, with $|\arg z|<\pi$ and $a>0$, we obtain for large n the limit value $\langle H\rangle\to-1$, according to (15) for $c=1$. A similar analysis of the Laguerre–Selberg integral (for $w=0$) was performed in [10], but it was restricted to the computation of $\langle H\rangle$ at leading order in n. Here we tackle the problem of the variance of H for square data matrices. From (49) we obtain

$$\mathrm{var}(H)=\langle H^2\rangle-\langle H\rangle^2=\frac{1}{(\beta n^2)^2}\{\widehat{P}''_n(0)-\widehat{P}'_n(0)^2\}=\frac{1}{n^2}\sum_{j=0}^{n-1}\psi_1\left(\frac{\beta}{2}(j+1)\right).\tag{51}$$

After a somewhat lengthy calculation, we managed to extract the large n asymptotics

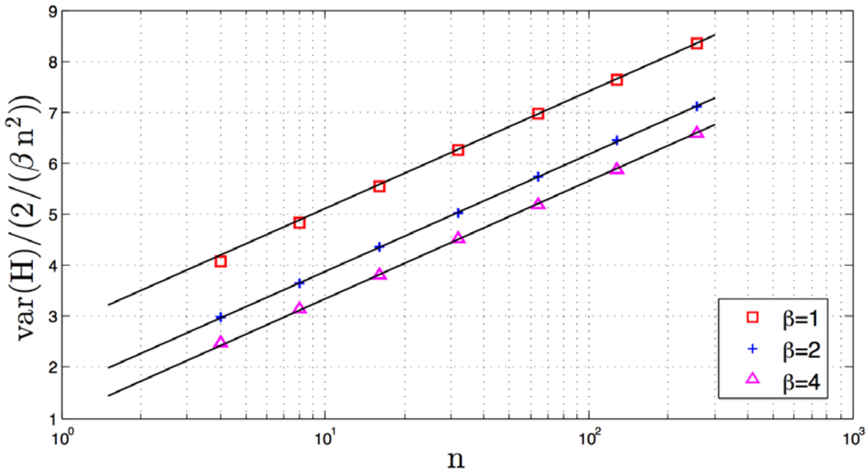

Figure 3. Rescaled variance of the log-determinant $H = n^{-1}\sum_i \ln(\lambda_i)$ for $c = 1$. Each point is produced sampling $N = 10^6$ Wishart matrices of size n for $\beta = 1, 2$ and 4. The error for each point is of order $\mathcal{O}(10^{-2})$, not visible in the picture. The solid lines are the exact result (52).

$$\frac{\mathrm{var}(H)}{\omega_\beta(n, 2)} = \ln n + \ln(\beta/2) - \psi_0(\beta/2) + 1 + K_\beta + o(1), \tag{52}$$

with a constant K_β given by

$$K_\beta = (2/\beta^2) \int_0^\infty t\, \mathrm{e}^{-2t(1-1/\beta)}(\beta \mathrm{e}^{2t/\beta} - 2\mathrm{e}^t + 2 - \beta)(1 - \mathrm{e}^{-t})^{-1}(\mathrm{e}^{2t/\beta} - 1)^{-2}\mathrm{d}t. \tag{53}$$

Some special values are $K_1 = \pi^2/8 - \ln 2$, $K_2 = 0$, $K_4 = 1 - \pi^2/8$. This result has been verified numerically, see figure 3. Note the logarithmic growth of (52) with n, in contrast to the $\mathcal{O}(1)$ limiting behavior of $\mathrm{var}(H)/\omega_\beta(n, 2)$ for $c > 1$. Such a logarithmic divergent variance is customary for *discontinuous* spectral linear statistics in RMT, the paradigmatic example being the number variance [20, 34, 36, 39]. Notice that the function $\ln \lambda$ is indeed discontinuous at $\lambda = 0$. However, as long as $s < s_c$ the support of the equilibrium measure $\rho_{s,w}(\lambda)$ does not contain the origin and this singularity is ineffective; only for $s = s_c$ we have $\lambda_-(s_c, w) = 0$, and at that point the singularity of $\ln \lambda$ starts being felt. The central limit theorem with logarithmically divergent variance for H has been proved in [41, 42] for $\beta = 1$. The subleading corrections to $\mathrm{var}(H)$ in (52) are instead a new result.

4.2. The statistics of atypically small log-determinants

We have claimed earlier that the rate function $\Psi_H(h)$ of H can be computed as the inverse Legendre–Fenchel transform of $J_H(s)$ only for $h > -1$. Why is this the case? As a matter of fact, the Gärtner–Ellis theorem has two hypotheses: first, the cumulant GF is required to be differentiable in the interior of its domain; second, the derivatives of the cumulant GF should diverge on the boundaries of the domain (a condition known as *steepness* [21, 51]). In our case, $J_H(s)$ is differentiable for all $s < s_c$ but the left derivative attains a *finite* value at the boundary point s_c: $\partial_s J_H(s) \to -1$ as $s \to s_c^-$. Hence, only a *local* version of the Gärtner–Ellis theorem holds and $\mathcal{P}_H(h) \approx \exp(-\beta n^2 \Psi_H(h))$ with $\Psi_H(h)$ being the branch of the Legendre–Fenchel transform of $J_H(s)$ for $h > -1$.

The 2D Coulomb gas analogy provides a rather intuitive physical picture of this obstruction. We have seen that the saddle-point density $\rho_{s,w}$ is bounded as long as $s < s_c$ (see (41) and (42)). When $s = s_c$, the lower edge of the saddle-point density reaches the origin $\lambda_-(s_c, w) = 0$ and $\rho_{s_c,w}(\lambda) \sim \lambda^{-1/2}$ acquires an inverse square-root singularity there. For $s > s_c$, the logarithmic part of the effective potential $V(\lambda) + U_{s,w}(\lambda)$ becomes *attractive*, giving rise to an electrostatic instability of the gas. As already discussed, $\rho_{s,w}(\lambda)$ is the typical distribution of the eigenvalues of $\mathcal{S}$ yielding a prescribed value of $H = \int \mathrm{d}\rho_{s,w}(\lambda) \ln \lambda$. Setting $w = \alpha = 0$ to simplify the discussion, we see that $\int \mathrm{d}\rho_{s,0}(\lambda) \ln \lambda > -1$ as long as $s < s_c$ and the critical value $H_{\mathrm{cr}} = -1$ corresponds to the critical density $\rho_{s_c,0}(\lambda) = \frac{1}{2\pi}\sqrt{(4-\lambda)/\lambda}\ \mathbf{1}_{\lambda \in (0,4)}$ obeying $\int \mathrm{d}\rho_{s_c,0}(\lambda) \ln \lambda = H_{\mathrm{cr}}$. A solution to the problem of smaller log-determinant $H < H_{\mathrm{cr}}$ would be achieved if the typical distribution of the eigenvalues corresponding to this anomalously small H were known.

What is then the behavior of the Coulomb gas constrained to have $H < H_{\mathrm{cr}}$? As suggested in [16], a failure of the steepness condition may be the hallmark of split-off phenomena of random variables. Guided by numerics and intuition, since the function $\ln \lambda$ is divergent for $\lambda \downarrow 0$, we expect that atypically small values of $H = n^{-1}\sum_{i=1}^{n} \ln \lambda_i < H_{\mathrm{cr}}$ are driven by the statistical behavior of the smallest eigenvalue $\lambda_{\min}$. For $H > H_{\mathrm{cr}}$ the Coulomb gas particles behave 'cooperatively' to accommodate atypical values of H (each of the random variables $\ln \lambda_i$'s contributes to realize H). On the contrary, large fluctuations of $H < H_{\mathrm{cr}}$ are typically realized by fluctuations of $\lambda_{\min}$ to the left (the random variable $\ln \lambda_{\min}$ contributes macroscopically to H). This line of reasoning would imply a change of scaling (speed) in the large n behavior of the probability density of H. The idea is to split the contribution of the Coulomb gas to H in two parts:

$$H[\rho] = \frac{1}{n}\ln \lambda_{\min} + H[\tilde{\rho}], \quad \text{where} \quad \tilde{\rho}(\lambda) = \frac{1}{n}\sum_{i:\lambda_i \neq \lambda_{\min}} \delta(\lambda - \lambda_i). \tag{54}$$

The probability density of H can be written as

$$\mathcal{P}_H(h) = \int_0^{+\infty} \mathrm{d}x \mathcal{P}_H(h \mid \lambda_{\min} = x)\mathcal{P}_{\lambda_{\min}}(x). \tag{55}$$

At this point we need to understand the distribution of the smallest eigenvalue $\mathcal{P}_{\lambda_{\min}}(x)$ and the conditional probability $\mathcal{P}_H(h \mid \lambda_{\min} = x)$. It is easy to show that the probability density function of the smallest eigenvalue behaves for large n as

$$\mathcal{P}_{\lambda_{\min}}(x) \approx \mathrm{e}^{-\beta n^2 x/2}, \tag{56}$$

corresponding to a typical value $\langle \lambda_{\min} \rangle = 2/(\beta n^2)$ and $\mathrm{var}(\lambda_{\min}) = 4/(\beta^2 n^4)$ at leading order in n. Typical fluctuations of order $\mathcal{O}(n^{-2})$ to the right of $\langle \lambda_{\min} \rangle$ are irrelevant for the statistical behavior of H. On the contrary the typical fluctuations to the left ($\lambda_{\min} < \langle \lambda_{\min} \rangle$) play a significant role due to the divergent character of $\ln \lambda_{\min}$ for $\lambda_{\min} \downarrow 0$. Roughly speaking, the *typical* fluctuations of order $\mathcal{O}(n^{-2})$ of the smallest eigenvalues do not change the limiting macroscopic density of the eigenvalues ($\tilde{\rho}(\lambda) \simeq \rho_{s_c,0}(\lambda)$,

J. Stat. Mech. (2016) 043306

irrespective of the value of $H \leqslant H_{\rm crit}$), but nevertheless have dramatic consequences on the statistics of H.

Similar evaporation phenomena for both correlated and i.i.d. variables have been recently detected in a variety of contexts (see e.g. [7, 18, 23, 24, 48, 50, 53]). The new interesting twist here is that the split-off is realized by the *smallest* (and not the usual *largest*) of the random variables involved. Here, using $H[\tilde{\rho}] \simeq H[\rho_{s_c,0}] = -1$, for $0 \leqslant x \leqslant \langle \lambda_{\rm min} \rangle$ we have:

$$\mathcal{P}_H(h|\lambda_{\rm min} = x) = \mathcal{P}_H\left(h = \frac{1}{n}\ln\lambda_{\rm min} + H[\tilde{\rho}] \,\Big|\, \lambda_{\rm min} = x\right) \to \delta\left(h - \left(\frac{1}{n}\ln\lambda_{\rm min} - 1\right)\right). \tag{57}$$

Using the above result, from (55), we easily get

$$\mathcal{P}_H(h) \approx \mathrm{e}^{-n\tilde{\psi}_H(h)}, \quad \text{with} \quad \tilde{\psi}_H(h) = -1 - h, \tag{58}$$

for $h < -1$ (note the speed n in contrast to the speed n^2 in the 'democratic' Coulomb gas setting).

A further argument in support of this change of speed can be obtained for $\beta = 2$ and $\alpha = 0$ using a finite-n approach based on the Laguerre–Selberg integral (46). It is convenient to work directly at the level of the determinant of $\mathcal{S}$. Let $\mathcal{P}_{\hat{G}}(\hat{g})$ be the probability density of the determinant $\hat{G} = \det(\mathcal{S}) = \prod_{i=1}^n \lambda_i$ (without the power $1/n$). Its Mellin transform is given by

$$\hat{M}(\hat{s}) = \int_0^\infty \mathrm{d}\hat{g}\,\mathcal{P}_{\hat{G}}(\hat{g})\hat{g}^{\hat{s}-1} = \left(\frac{1}{n}\right)^{n(\hat{s}-1)} \frac{G(n+\hat{s})\Gamma(\hat{s})}{G(n+1)G(1+\hat{s})}, \tag{59}$$

where $G(x)$ is the Barnes G-function. Using the asymptotics [1]

$$\frac{G(n+\hat{s})}{G(n+1)} \sim \exp\left[n(1-\hat{s}) - n\ln n + \hat{s}\; n\ln n + \ln n[1/2 - \hat{s} + \hat{s}^2/2] - \ln(2\pi)(1-\hat{s})/2 + o(1/n)\right], \tag{60}$$

valid for $n \to \infty$, we obtain (with logarithmic accuracy)

$$\hat{M}(\hat{s}) \approx \mathrm{e}^{-n(\hat{s}-1)}. \tag{61}$$

This Mellin transform can be written also in terms of the probability density $\mathcal{P}_H(h)$ of $H = n^{-1}\ln\hat{G}$. Assuming $\mathcal{P}_H(h) \approx \mathrm{e}^{-n\tilde{\psi}_H(h)}$ for $h < -1$ we have

$$\hat{M}(\hat{s}) = \int_{-\infty}^\infty \mathrm{d}h\mathcal{P}_H(h)\mathrm{e}^{n(\hat{s}-1)h} \approx \int_{-\infty}^{-1} \mathrm{d}h\; \mathrm{e}^{-n[\tilde{\psi}_H(h) - (\hat{s}-1)h]} \approx \exp\left[-n\min_h(\tilde{\psi}_H(h) - (\hat{s}-1)h)\right]. \tag{62}$$

Here the integral has been truncated at $h = -1$ since $\mathcal{P}_H(h)$ decays faster (exponentially with speed n^2) for $h > -1$, and Laplace's approximation has been used in the last step. Matching (61) with (62), we eventually obtain $\tilde{\psi}_H(h)$ as in (58).

5. Conclusions

In summary, we have considered the joint statistics (including large deviation tails) of generalized and total variance of a large $n \times m$ Gaussian dataset. These observables are just the scaled log-determinant H and the trace T of the corresponding $n \times n$ covariance matrix. We have employed a powerful combination of two techniques: the standard Coulomb gas analogy of statistical physics, which allowed us to represent the eigenvalues of the covariance matrix as an interacting gas of charged particles, whose excess free energy is the cumulant generating function for our observables in the limit $n, m \to \infty$ with $c = m/n$ fixed, and a finite n, m approach based on the Laguerre–Selberg integral. Combining these two approaches, we complemented the Coulomb gas method with two interesting cases that fell out of its domain: (i) the case $c = 1$ (square datasets), for which the excess free energy is non-analytic in zero. This has the consequence that the variance of H grows logarithmically with n, with a subleading constant term that we could precisely characterize, and (ii) atypically small log-determinants, for which the corresponding rate function in Laplace space is non-steep. This implies an abrupt change of speed in the corresponding large deviation principle, which can be ascribed to the *split-off* of the smallest eigenvalue from the 'unperturbed' Marčenko–Pastur distribution. This picture is supported by numerical simulations and a saddle-point argument based on a finite n, m formula (see section 4.2).

It would be interesting to investigate whether our results could be extended to non-Gaussian and possibly correlated data matrices. Our derivation strongly relied on the data being normally distributed and a different approach seems to be needed for more general covariance matrices.

Acknowledgments

FDC acknowledges support by EPSRC Grant number EP/L010305/1 and partial support of Gruppo Nazionale di Fisica Matematica GNFM-INdAM. PV acknowledges the stimulating research environment provided by the EPSRC Centre for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, EP/L015854/1). FDC acknowledges hospitality by the LPTMS (Univ. Paris Sud) where this work was initiated. The results presented in this paper are part of the first author's *PhD Thesis* at the University of Bari, Italy. The authors thank P Facchi, A Lerario and M Marsili for helpful advices and very stimulating discussions. FDC thanks S Di Martino for technical support.

References

[1] Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (New York: Dover) (10th printing)

[2] Anderson T W 1984 *An Introduction to Multivariate Statistical Analysis* (*Wiley Series in Probability and Statistics* vol 114) (New York: Wiley)

[3] Anderson G W, Guionnet A and Zeitouni O 2009 *An Introduction to Random Matrices* 3rd edn (Cambridge: Cambridge University Press)

[4] Bai Z B and Silverstein J W 2004 CLT for linear spectral statistics of large-dimensional sample covariance matrices *Ann. Probab.* **32** 1A

[5] Barlett M S 1933 On the theory of statistical regression *Proc. R. Soc. Edinburgh.* **53** 260–83
[6] Beenakker C W J 1994 Universality of Brézin and Zee's spectral correlator *Nucl. Phys.* B **422** 515
[7] Bogacz L, Burda Z, Janke W and Wacław B 2007 Balls-in-boxes condensation on networks *Chaos* **17** 026112
[8] Brézin E, Itzykson C, Parisi G and Zuber J B 1978 Planar diagrams *Commun. Math. Phys.* **59** 35
[9] Chen Y and Manning S M 1994 Asymptotic level spacing of the Laguerre ensemble: a Coulomb fluid approach *J. Phys. A: Math. Theor.* **27** 3615–20
Chen Y and Manning S M 1996 Some eigenvalue distribution functions of the Laguerre ensemble *J. Phys. A: Math. Gen.* **29** 7561–79
[10] Cicuta G M and Mehta M L 2000 Probability density of determinants of random matrices *J. Phys. A: Math. Gen.* **33** 8029
[11] Colomo F and Pronko A G 2013 Third-order phase transition in random tilings *Phys. Rev.* E **88** 042125
[12] Cunden F D and Vivo P 2014 Universal covariance formula for linear statistics on random matrices *Phys. Rev. Lett.* **113** 070202
[13] Cunden F D, Facchi P and Vivo P 2015 Joint statistics of quantum transport in chaotic cavities *Europhys. Lett.* **110** 50002
[14] Cunden F D, Maltsev A and Mezzadri F 2015 Fluctuations in the 2D one-component plasma and associated fourth-order phase transition *Phys. Rev.* E **91** 060105
[15] Cunden F D, Mezzadri F and Vivo P 2015 A unified fluctuation formula for one-cut β-ensembles of random matrices *J. Phys. A: Math. Theor.* **48** 315204
[16] Cunden F D, Facchi P and Vivo P 2016 A shortcut through the Coulomb gas method for spectral linear statistics on random matrices *J. Phys. A: Math. Theor.* **49** 135202
[17] Dean D S and Majumdar S N 2006 Large deviations of extreme eigenvalues of random matrices *Phys. Rev. Lett.* **97** 160201
Dean D S and Majumdar S N 2008 Extreme value statistics of eigenvalues of Gaussian random matrices *Phys. Rev.* E **77** 041108
[18] De Pasquale A, Facchi P, Parisi G, Pascazio S and Scardicchio A 2010 Phase transitions and metastability in the distribution of the bipartite entanglement of a large quantum system *Phys. Rev.* A **81** 052324
[19] Dumitriu I and Edelman A 2002 Matrix models for beta ensembles *J. Math. Phys.* **43** 5830
[20] Dyson F J 1962 Statistical theory of the energy levels of complex systems *J. Math. Phys.* **3** 140, 157, 166, 1191, 1199
[21] Ellis R S 1984 Large deviations for a general class of random vectors *Ann. Probab.* **12** 1–2
Ellis R S 1999 The theory of large deviations: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence *Physica* D **133** 106–36
[22] Facchi P, Marzolino U, Parisi G, Pascazio S and Scardicchio A 2008 Phase transitions of bipartite entanglement *Phys. Rev. Lett.* **101** 050502
[23] Facchi P, Florio G, Parisi G, Pascazio S and Yuasa K 2013 Entropy-driven phase transitions of entanglement *Phys. Rev.* A **87** 052324
[24] Filiasi M, Livan G, Marsili M, Peressi M, Vesselli E and Zarinelli E 2014 On the concentration of large deviations for fat tailed distributions, with application to financial data *J. Stat. Mech.* P09030
[25] Fisher R A 1915 Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population *Biometrika* **10** 507–21
[26] Forrester P J and Warnaar S O 2008 The importance of the Selberg integral *Bull. Am. Math. Soc.* **45** 489
[27] Gärtner J 1977 On large deviations from the invariant measure *Theory Probab. Appl.* **22** 24–39
[28] Grabsch A and Texier C 2015 Capacitance and charge relaxation resistance of chaotic cavities joint distribution of two linear statistics in the Laguerre ensemble of random matrices *Europhys. Lett.* **109** 50004
Grabsch A and Texier C 2016 Distribution of spectral linear statistics on random matrices beyond the large deviation function—Wigner time delay in multichannel disordered wires (arXiv:1602.03370) (preprint)
[29] Jiang D, Jiang T and Yang F 2012 Likelihood ratio tests for covariance matrices of high-dimensional normal distributions *J. Stat. Plan. Inference* **142** 2241
[30] Johnsson R A and Wichern D W 2007 *Applied Multivariate Statistical Analysis* (Englewood Cliffs, NJ: Prentice Hall)
[31] Johnstone I M 2006 High dimensional statistical inference and random matrices *Proc. Int. Congr. Math.* (arXiv:math/0611589 [math.ST]) (preprint)
[32] Jonsson D 1982 Some limit theorems for the eigenvalues of a sample covariance matrix *J. Multivariate Anal.* **12** 1
[33] Le Caër G and Delannay R 2007 The fixed-trace β-Hermite ensemble of random matrices and the low temperature distribution of the determinant of an $N \times N$ β-Hermite matrix *J. Phys. A: Math. Theor.* **40** 1561

[34] Majumdar S N, Nadal C, Scardicchio A and Vivo P 2009 Index distribution of Gaussian random matrices *Phys. Rev. Lett.* **103** 220603

Majumdar S N, Nadal C, Scardicchio A and Vivo P 2009 How many eigenvalues of a Gaussian random matrix are positive? *Phys. Rev.* E **83** 041105

[35] Majumdar S N and Vivo P 2012 Number of relevant directions in principal component analysis and wishart random matrices *Phys. Rev. Lett.* **108** 200601

[36] Majumdar S N, Schehr G, Villamaina D and Vivo P 2013 Large deviations of the top eigenvalue of large Cauchy random matrices *J. Phys. A: Math. Theor.* **46** 022001

[37] Majumdar S N and Schehr G 2014 Top eigenvalue of a random matrix: large deviations and third order phase transition *J. Stat. Mech.* P01012

[38] Marčenko V A and Pastur L A 1967 Distribution of eigenvalues for some sets of random matrices *Math. USSR—Sb.* **1** 457

[39] Marino R, Majumdar S N, Schehr G and Vivo P 2016 Number statistics for β-ensembles of random matrices: applications to trapped fermions at zero temperature (arXiv:1601.03178) (preprint)

[40] Muirhead R J 2005 *Aspect of Multivariate Statistical Theory* (New York: Wiley)

[41] Nguyen H H and Ve V 2014 Random matrices: law of the determinant *Ann. Probab.* **42** 146–67

[42] Rempał G and Wesołowski J 2005 Asymptotics for products of independent sums with an application to Wishart determinants *Stat. Probab. Lett.* **74** 129–38

[43] Rouault A 2005 Pathwise asymptotic behavior of random determinants in the uniform Gram and Wishart ensembles (arXiv:math/0509021) (preprint)

Rouault A 2007 Asymptotic behavior of random determinants in the Laguerre, Gram and Jacobi ensembles *Alea* **3** 181–230

[44] Saff E B and Totik V 1991 *Logarithmic Potentials with External Fields* (Berlin: Springer)

[45] Selberg A 1942 On the zeros of Riemann's zeta-function *Skr. Norske Vid. Akad. Oslo I.* **10** 1

[46] Serfaty S 2014 *Coulomb Gases and Ginzburg-Landau Vortices* (New York: Courant Institute of Mathematical Sciences)

Serfaty S 2014 Ginzburg-Landau vortices, Coulomb gases, and renormalized energies *J. Stat. Phys.* **154** 660–80

[47] Silverstein J W 1985 The smallest eigenvalue of a large dimensional Wishart matrix *Ann. Probab.* **13** 1364–8

[48] Szavits-Nossan J, Evans M R and Majumdar S N 2014 Constraint-driven condensation in large fluctuations of linear statistics *Phys. Rev. Lett.* **112** 020602

[49] Tao T and Vu V 2012 A central limit theorem for the determinant of a Wigner matrix *Adv. Math.* **231** 74–101

[50] Texier C and Majumdar S N 2013 Wigner time-delay distribution in chaotic cavities and freezing transition *Phys. Rev. Lett.* **110** 250602

Texier C and Majumdar S N 2014 Erratum: Wigner time-delay distribution in chaotic cavities and freezing transition [*Phys. Rev. Lett.* 110, 250602 (2013) *Phys. Rev. Lett.* **112** 139902

[51] Touchette H 2009 The large deviation approach to statistical mechanics *Phys. Rep.* **478** 1

[52] Tricomi F G 1957 *Integral Equations* (*Pure and Applied Mathematics*) (London: Interscience)

[53] Vivo P, Majumdar S N and Bohigas O 2007 Large deviations of the maximum eigenvalue in Wishart random matrices *J. Phys. A: Math. Theor.* **40** 4317–37

Vivo P, Majumdar S N and Bohigas O 2008 Distributions of conductance and shot noise and associated phase transitions *Phys. Rev. Lett.* **101** 216809

Vivo P, Majumdar S N and Bohigas O 2010 Probability distributions of linear statistics in chaotic cavities and associated phase transitions *Phys. Rev.* B **81** 104202

[54] Wigner E P 1957 *Statistical Properties of Real Symmetric Matrices with Many Dimensions* (*Canadian Mathematical Congress Proc.*) (Toronto: University of Toronto Press) pp 174

[55] Wishart J 1928 The generalised product moment distribution in samples from a normal multivariate population *Biometrika* **20A** 32